

# Review of: "A Simple Preprocessing Method Enhances Machine Learning Application to EEG Data for Differential Diagnosis of Autism"

Norizam Sulaiman<sup>1</sup>

<sup>1</sup> University Malaysia Pahang

Potential competing interests: No potential competing interests to declare.

Overall, the contents of the paper are good. However, they need some improvement in certain areas in order to enhance the quality of the manuscript, as stated below:

Abstract section:

1. Authors cannot start this section with the methodology. This section is supposed to be started with the background or motivation of the study.
2. The objective of the study is not clearly written.
3. Authors need to explain what is meant by 38 figures in the last paragraph of this section.

Graphical Abstract section:

Satisfactory. Just might need to change the accuracy (94%) to be the same as stated in the Abstract and Results sections.

Introduction section:

1. It seems that there is not enough literature review in this section. Only 7 references are cited in this section, which is not enough. It is suggested that the authors cite more than 20 references in order to provide a good comprehensive review of the topic being studied.
2. There is no literature review on machine learning, especially on the KNN classifier.
3. It is good that the objective of the study is clearly described in this section.
4. Authors need to provide a comprehensive literature review on the mathematically topological approach and might need to cite references related to this term and define this term in detail.
5. Authors need to provide a citation for the MST algorithm (page 3).

Patients and Methods sub-section:

It is more preferable if this sub-section is relocated under the Methods section.

Methods section:

1. It is suggested that authors start this section with a process flowchart in order to show the overall processes and parameters that are involved in the study.
2. It is more preferable if authors can share the diagram of the EEG electrode montage to show the location of the EEG measurement area beside the EEG channels' positions.
3. Authors are required to specify the sampling frequency employed by the EEG instrument during EEG measurement.
4. Authors are required to provide an equation or formula for calculating the EEG Manhattan Distance Matrix and Minimum Spanning Tree (MST).
5. Authors need to explain Figures 2 and 3 in the text paragraph.
6. Authors need to explain how the 38 numbers were produced in the 2nd paragraph on page 7.
7. Authors need to explain the test and training ratio that was used in the study.
8. Authors need to explain the machine learning that was used in the study, including the confusion matrix.
9. The explanation of Figure 4 is too short. Authors need to explain how the electrode numbers and number of links in the MST can be employed to train the classifier.
10. Authors need to explain the cross-validation procedure that was employed in the study.

Predictive Modeling sub-section:

1. Author needs to explain the use of cross-validation in the study. How many folds should be used?
2. Author needs to explain the configuration parameters for the KNN algorithm.

Training and Testing Protocol sub-section:

The steps under this section are clearly described.

Natural Cluster sub-section:

The literature review for the PST unsupervised machine learning system should be done and put in the Introduction section.

Results section:

1. The results of the study shown in Table 1 are good. However, since the machine learning algorithm, KNN, is employed in the study, the results also should come with the confusion matrix.
2. The results for cross-validation are not shared in this section.
3. Author needs to explain Figure 5 in more detail. Author needs to explain how the PST can produce what is shown in Figure 5.

Discussion section:

The content of this section is good. However, the author might need to emphasize the strengths and weaknesses of the study.

Statement and Declaration section:

Acceptable.

Other References section:

The total number of references is 17. What is the purpose of these other references? Can they be combined with the references under the References section?

References section:

1. Total number of references is 9. The total number of main references is 9, which is not enough for the study.
2. It is recommended that the author provide more recent references for the study, such as publications from 2020-2023.